PCT/EP2004/006805

Europäisches Patentamt
European Patent Office
Office européen des brevets

Bescheinigung Certificate Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

03394056.0

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Europäisches Patentamt | European Patent Office | Office européen des brevets

Anmeldung Nr:
Application no.: 03394056.0
Demande no:

Anmeldetag:
Date of filing: 20.06.03
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

NESTEC S.A.
Avenue Nestlé 55
CH-1800 Vevey
SUISSE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Modulation of coffee flavour precursor levels in green coffee grains

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12N9/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL
PT RO SE SI SK TR LI

Modulation of Coffee Flavour Precursor Levels in Green Coffee Grains

Background of the Invention:

Coffee contains a highly complex mixture of flavour molecules. Extensive research on the composition of instant and fresh ground coffee beverages has, to date, identified more than 850 compounds, many of which are flavour active molecules (Flament, I (2002) Coffee Flavor Chemistry, John Wiley and Sons, UK). However, few of the final coffee flavour molecules found in the cup are present in the raw material, the green grain (green beans) of the plant species *Coffea arabica* or *Coffea canephora (robusta)*. In fact, the majority of the coffee flavour compounds are generated during one or more of the multiple processing steps that occur from the harvest of the ripe red coffee cherries to the final roasted ground coffee product, or extracts thereof, for example soluble coffee products.

The various steps in the production of coffee are described in Smith, A.W., in Coffee; Volume 1: Chemistry pp 1-41, Clark, R.J. and Macrea, R. eds, Elsevier Applied Science London and New York, 1985; Clarke, R.J., in Coffee: Botany, Biochemistry, and Production of Beans and Beverage, pp 230-250 and pp 375-393; and Clifford, M.N. and Willson, K.C. eds, Croom Helm Ltd, London. Briefly, the process starts with the collection of mature, ripe red cherries. The outer layer, or pericarp, can then be removed using either the dry or wet process. The dry process is the simplest and involves 1) classification and washing of the cherries, 2) drying the cherries after grading (either air drying or mechanical drying), and 3) dehusking the dried cherries to remove the dried pericarp. The wet process is slightly more complicated, and generally leads to the production of higher quality green beans. The wet process is more often associated with *C. arabica* cherries. The wet process consists of 1) classification of the cherries, 2) pulping of the cherries, this step is done soon after harvest and generally involves mechanical removal of the "pulp", or pericarp, of the mature cherries, 3) "fermentation", the mucilage that remains attached to the grain of the cherries after pulping is removed by allowing the grain plus attached mucilage to be incubated with water in tanks using a batch process. The "fermentation" process is allowed to continue up to 80 hours, although often 24 hours is generally enough to allow an acceptable fermentation and to cause the pH to drop from around 6.8-6.9 to 4.2-4.6, due to various enzymatic activities and the metabolic action of microorganisms which grow during the fermentation, 4) drying, this

step involves either air or mechanical hot air drying of the fermented coffee grain and 5) "hulling", this step involves the mechanical removal of the "parch" of the dried coffee grain (dried parchment coffee) and often the silverskin is also removed at this stage. After wet or dry processing, the resulting green coffee grain are often sorted, with most sorting procedures being based on grain size and/or shape.

The next step in coffee processing is the roasting of the green grain after dehusking or dehulling of dry or wet processed coffee, respectively. This is a time-dependent process which induces significant chemical changes in the bean. The first phase of roasting occurs when the supplied heat drives out the remaining water in the grain. When the bulk of the water is gone, roasting proper starts as the temperature rises towards 190-200°C. The degree of roasting, which is usually monitored by the colour development of the beans, plays a major role in determining the flavour characteristics of the final beverage product. Thus, the time and temperature of the roasting are tightly controlled in order to achieve the desired coffee flavour profile. After roasting, the coffee is ground to facilitate extraction during the production of the coffee beverage or coffee extracts (the latter to be used to produce instant coffee products). Again, the type of grinding can influence the final flavour of the beverage.

While a considerable amount of research has been carried out on the identification of the flavour molecules in coffee, much less work has been done regarding the physical and chemical reactions which occur within the coffee grains during each of the processing steps. This latter point is particularly evident for the roasting reaction, where the large number of grain constituents undergo an extremely complex series of heat induced reactions (Homma, S. 2001, In "Coffee: Recent Developments". R.J. Clarke and O.G. Vitzthum eds, Blackwell Science, London; Yeretzian, C., *et al* ((2002) Eur. Food Res. Technol. 214, 92-104; Flament, I (2002) Coffee Flavor Chemistry, John Wiley and Sons, UK; Reineccius, G.A., "The Maillard Reaction and Coffee Flavor" Conference Proceedings of ASIC, 16th Colloque, Kyoto, Japan 1995). Because of the large number of the potential reactants, and the complexity of the reactions involved, little research data currently exist which establishes strong links between specific biological molecules present in the green beans and the numerous flavour molecules identified in a cup of coffee.

Nonetheless, while the details of most of the reactions that occur during the different steps of coffee processing remain relatively unclear, it is thought that an important flavour generating

reaction responsible for many of the flavours associated with coffee aroma is the "Maillard" reaction during coffee roasting. A vigorous Maillard reaction occurs between the grain reducing sugars/polysaccharide degradation products and the amino group containing molecules (particularly the proteins, peptides, and amino acids) during the roasting step.

Because the Maillard reaction apparently makes an important contribution to the generation of coffee flavour and aroma molecules during coffee roasting, there might be an association between the levels of primary Maillard reactants in the green beans and the quality of the flavour/aroma developed after roasting.

As noted above, an important group of substrates in the Maillard reaction are the amino acids, peptides and proteins. To date, there are no firm associations between the levels, or types, of amino group containing molecules in different coffees and the qualities of those coffees. Although the amino acid contents of *arabica* and *robusta* coffee beans have been analysed (M.N. Clifford 1985, In "Coffee: Botany, Biochemistry, and Production of Beans and Beverage", Clifford and Willson eds, Croom Helm London; Arnold *et al* 1996, Z. Lebensm Unters Forsch., Vol 199, 22-25; Ludwig *et al* 2000, Eur. Food Res Technol., Vol 211, 111-116.), there are no reports directly linking specific levels or ratios of amino acids and high or low flavour qualities. Using 2-D electrophoresis, it has been shown that differences exist in the levels and amounts of the major storage proteins in *arabica* and *robusta* green coffee beans – however, no association between these storage protein differences and flavour quality was noted (Rogers *et al*, 1999, Plant Physiol. Biochem. Vol 37, 261-272). It has also recently been found that small differences exist between the storage proteins of immature and mature coffee beans, which have different flavour qualities (Montavon, P. *et al*, 2003, J. Agric and Food Chemistry Vol 51, 2328-2334). However, because there are many changes occurring during seed maturation, this latter work only suggests a link may exist between the quality improvement caused by maturation and the differences seen in the 2-D gel patterns of the main coffee storage proteins. Overall, currently no clear evidence exists linking any differences seen for the coffee storage proteins, or other major green bean proteins, and the flavour qualities of coffee.

It has recently been shown that there are differences in the profiles of peptides isolated from *arabica* and *robusta* green beans (Ludwig *et al* 2000, Eur. Food Res Technol., Vol 211, 111-116.). Although their results showed that the *arabica* and *robusta* peptide extracts differ in

their aroma precursor profile, the data presented in this report do not identify which component(s) in the extracts is / are responsible for these aroma profile differences. These workers also detected at least two different proteinase activities in crude extracts of the green coffee, but they did not correlate any specific activities with aroma/flavour quality (Ludwig *et al* 2000, Eur. Food Res Technol., Vol 211, 111-116). Finally, it is also thought that the very high temperatures used during the later stages of green coffee grain roasting cause substantial cleavage of the proteins present in the coffee grain (Homma, S. 2001, In "Coffee: Recent Developments". R.J. Clarke and O.G. Vitzthum eds, Blackwell Science, London; Montavon, P., *et al* 2003, "Changes in green coffee protein profiles during roasting", J. Agric. Food Chem. 51, 2335-2343). However, the overall scheme for this protein degradation is very poorly understood, but presumably depends on, among other things, the precise state of the main coffee proteins in the raw material before the start of roasting. To our knowledge, there are no other significant reports addressing the possibility that peptide profiles in coffee could be involved in the production of coffee aroma/flavour.

In the roasting of the fermented seeds of *Theobroma cacao* (cocoa beans), there would appear to be an involvement of seed amino acids and peptides in the development of Maillard reaction aromas/flavours. Relative to other seeds, *T. cacao* seeds have been shown to have an unusually high level of aspartic proteinase activity (Biehl, B., Voigt, J., Voigt, G., Heinrichs, H., Senyuk, V. and Bytof, G. (1994) "pH dependent enzymatic formation of oligopeptides and amino acids, the aroma precursors in raw cocoa beans". In *The Proceedings of the 11th International Cocoa Research Conference,* 18-24 July 1993, Yamoussoukro, Ivory Coast). In order to produce cocoa beans with a high level of cocoa flavour precursors, it is necessary to carry out a natural fermentation step (unfermented beans develop little flavour when roasted). During this fermentation step, the sugars in the pulp are fermented, generating high levels of acids, particularly acetic acid (Carr, J.G. (1982) Cocoa. In *Fermented Foods.* Economic Microbiology. Vol 7. pages 275-292. (A.H. Rose ed). Academic Press). As the fermentation continues, the pH in the seed decreases and the cell structure becomes disrupted. The low pH triggers the abundant cacao seed aspartic proteinase to become mobilized and/or activated, resulting in a massive degradation of cellular protein (Biehl, B., Passern, D., and Sagemann, W. (1982) "Effect of Acetic Acid on Subcellular Structures of Cocoa Bean Cotylydons". J. Sci. Food Agric. 33, 1101-1109; Biehl., B., Brunner, E., Passern, D., Quesnel, V.C., and Adomako, D. (1985) "Acidification, proteolysis and flavour potential in fermenting cocoa beans". J. Sci. Food Agric. 36, 583-598). Peptides and amino acids have

been shown to be cocoa flavour precursors (Rohan, T. (1964) "The precursors of chocolate aroma: a comparative study of fermented and unfermented cocoa beans". J. Food Sci., 29, 456-459; Voigt, J. and Biehl, B. (1995) "Precursors of the cocoa specific aroma components are derived from the vicilin-class (7S) globulin of the cocoa seeds by proteolytic processing". Bot. Acta 108, 283-289). Thus, the *T. cacao* seed aspartic proteinase, together with a seed serine carboxypeptidase, have been proposed to be critical for the generation of cocoa flavour precursors during fermentation (Voigt, J. and Biehl, B. (1995) "Precursors of the cocoa specific aroma components are derived from the vicilin-class (7S) globulin of the cocoa seeds by proteolytic processing". Bot. Acta 108, 283-289; Voigt, J., Heinrichs, H., Voigt, G. and Biehl, B. (1994) "Cocoa-specific aroma precursors are generated by proteolytic digestion of the vicilin-like globulin of cocoa seeds". Food Chemistry, 50, 177-184.) The gene encoding the abundant cacao seed aspartic proteinase has been identified and a method to over-express this protein in cacao seeds which can generate increased levels of cacao flavour precursor amino acids and peptides in fermented cocoa beans has recently been described in International Patent Publication No. 02/04617, the whole contents of which are incorporated herein by reference. However, the teaching of International Patent Publication No. 02/04617 is directed towards cacao seeds, which undergo a specific long acid fermentation step, unlike coffee grains which do not.

Objects of the Invention

It is an object of the present invention to improve the flavour quality of coffee.

More specifically, it is an object of the present invention to improve the levels of the flavour precursors in the raw material (the green grain) so that, following post harvest treatment and roast-processing, an improved flavour is achieved. Without being bound by theory, it is believed that, if there are variations in the levels of peptides and protein degradation between coffees with significantly different flavours, then it is possible that these variations could be due to differences in the endogenous proteinase activities in these different grains. This difference might be detectable at the level of mRNA expression by variations in the levels of expression for particular seed proteinase genes.

Statements of the Invention

The present invention involves, therefore, identifying gene sequences encoding for coffee grain (seed) specific proteinases and showing that there are indeed variations in the expression of these genes in *arabica* and *robusta.*

More specifically, the present invention discloses a major coffee cysteine proteinase (CcCP-1), a major coffee cysteine proteinase inhibitor (CcCPI-1) and coffee aspartic proteinases (CcAP-1 and CcAP-2), all of which are expressed in coffee seeds. We further show how either over-expression of these proteins specifically late in seed development, or the reduced expression of these proteins during late seed development, can alter the amino acid/peptide/protein profile of the mature beans. By using one or more of the disclosed gene sequences and gene constructs to alter the amino acid/peptide/protein profile of the mature beans, we disclose a new method to alter the flavour precursor profile of mature coffee beans, and thus thereby allow the production of roasted coffee beans with altered flavours.

In a first aspect, the present invention provides an isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide having cysteine proteinase activity, wherein the amino acid sequence of the polypeptide and the amino acid sequence of SEQ ID No. 2 have at least 70%, preferably at least 80%, sequence identity based on the ClustalW alignment method; or the complement of the nucleotide sequence, wherein the complement contains the same number of nucleotides as the nucleotide sequence, and the complement and the nucleotide sequence are 100% complementary. Preferably, the amino acid sequence of the polypeptide and the amino acid sequence of SEQ ID No. 2 have at least 85%, preferably at least 90%, optionally at least 95%, sequence identity based on the ClustalW alignment method. Preferably, the nucleotide sequence comprises the nucleotide sequence of SEQ ID No. 1. Preferably, the polypeptide comprises the amino acid sequence of SEQ ID No. 2.

In a second aspect, there is provided an isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide having cysteine proteinase inhibitor activity, wherein the amino acid sequence of the polypeptide and the amino acid sequence of SEQ ID No. 4 have at least 70%, preferably at least 80%, sequence identity based on the ClustalW alignment method; or the complement of the nucleotide sequence, wherein the complement contains the same number of nucleotides as the nucleotide sequence, and the complement and the nucleotide sequence are 100% complementary. Preferably, the amino acid sequence of the polypeptide and the amino acid sequence of SEQ ID No. 4 have at least 85%, preferably at

least 90%, optionally at least 95%, sequence identity based on the ClustalW alignment method. Preferably, the nucleotide sequence comprises the nucleotide sequence of SEQ ID No. 3. Preferably, the polypeptide comprises the amino acid sequence of SEQ ID No. 4.

In a third aspect, there is provided an isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide having aspartic endoproteinase activity, wherein the amino acid sequence of the polypeptide and the amino acid sequence selected from SEQ ID No. 6 or 8, preferably SEQ ID No. 8, have at least 75%, preferably at least 80%, sequence identity based on the ClustalW alignment method, or the complement of the nucleotide sequence, wherein the complement contains the same number of nucleotides as the nucleotide sequence, and the complement and the nucleotide sequence are 100% complementary. Preferably, the amino acid sequence of the polypeptide and the amino acid sequence selected from SEQ ID No. 6 or 8, preferably SEQ ID No. 8, have at least 85%, preferably at least 90%, optionally at least 95%, sequence identity based on the ClustalW alignment method. Preferably, the nucleotide sequence comprises the nucleotide sequence of SEQ ID No. 5 or 7, preferably SEQ ID No. 7. Preferably, the polypeptide comprises the amino acid sequence of SEQ ID No. 6 or 8, preferably SEQ ID No. 8.

In a further aspect, there is provided a vector comprising the polynucleotide of any one of first to third aspects of the invention.

In a further aspect, there is provided a recombinant DNA construct comprising the polynucleotide of any one of first to third aspects of the invention, operably linked to a regulatory sequence.

In a further aspect, there is provided a method for transforming a cell comprising transforming the cell with the polynucleotide of any one of first to third aspects of the present invention.

In a further aspect, there is provided a cell comprising the aforementioned recombinant DNA construct, which cell is preferably a prokaryotic cell, an eukaryotic cell or a plant cell, preferably a coffee cell.

In a further aspect, there is provided a transgenic plant comprising such a transformed cell.

In the present application, coffee cherry terms are defined as follows: coffee cherry; entire fruit; exocarp, skin; pericarp, fleshy major outer layer of cherry; and grain, coffee seed. For a fuller explanation of thee terms, reference is made to Clarke, R.J., in Coffee: Botany, Biochemistry, and Production of Beans and Beverage, pp 230, Clifford, M.N. and Willson, K.C. eds, Croom Helm Ltd, London, the contents of which are incorporated in their entirety.

Brief Description of the Invention

The invention can be understood from the following detailed description and the accompanying Sequence Listing which forms part of the present application.

Table 1 hereunder lists the polypeptides that are described herein, along with the corresponding sequence identifier (SEQ ID No) as used in the attached listing.

Table 1:

SEQ ID No 1 (CcCP1 : Cysteine proteinase, nucleic acid and its corresponding amino acid)
SEQ ID No 2 (CcCP1: Cysteine proteinase, amino acid)
SEQ ID No 3 (CcCPI-1 : Cysteine proteinase Inhibitor, nucleic acid and its corresponding amino acid)
SEQ ID No 4 (CcCPI-1: Cysteine proteinase Inhibitor, amino acid)
SEQ ID No 5 (CcAP1 : Aspartic endoproteinase 1, nucleic acid and its corresponding amino acid)
SEQ ID No 6 (CcAP1 : Aspartic endoproteinase 1, amino acid)
SEQ ID No 7 (CcAP2 : Aspartic proteinase 2, nucleic acid and its corresponding amino acid)
SEQ ID No 8 (CcAP2 : Aspartic proteinase 2, amino acid)

The sequence listing employs the one letter codes for nucleotide sequence characters and the three letter codes for amino acids as defined for IUPAC-IUBMB Standards and as described in Nucleic Acids Research 13:3021-3030 (1985), which is incorporated herein by reference.

Drawings

In the drawings,

Figure 1 shows a Northern blot analysis of cysteine proteinase gene in different tissues of *Coffea arabica*, in which the lanes are labeled R: root, S: stem, L: young leaves; and SG, LG, Y and Red are grain from small green fruit, large green fruit, yellow fruit and red fruit, respectively. Five micrograms of total RNA was loaded in each lane. MW is an RNA size ladder. Panel A illustrates an autoradiography after 24 hours exposure and Panel B demonstrates the ethidium bromide staining of the gels prior to blotting;

Figure 2 shows a Northern blot analysis of Cysteine proteinase inhibitor gene in different tissues of *Coffea arabica*, in which the lanes are labeled R: root, S: stem, L: young leaves and SG, LG, Y and Red for grain from small green fruit, large green fruit, yellow fruit and red fruit, respectively. Five micrograms of total RNA was loaded in each lane. MW is an RNA size ladder. Panel A illustrates an autoradiography after 24 hours exposure and panel B demonstrates the ethidium bromide staining of the gels prior to blotting;

Figure 3 shows a Northern blot analysis of Cysteine proteinase inhibitor gene in different stages of development of *Coffea arabica* (ARA) and *Coffea robusta* (ROB) fruit. The lanes are labeled small green fruit (SG), large green fruit (LG), yellow fruit (Y) and red fruit (Red), respectively. Five micrograms of total RNA was loaded in each lane. MW is an RNA size ladder. Panel A illustrates an autoradiography after 24 hours exposure. Panel B demonstrates the ethidium bromide staining of the gels prior to blotting; and

Figure 4 shows a Northern blot analysis of aspartic proteinase 2 (CcAP2) gene in different tissues of *Coffea arabica*, in which the lanes are labelled R: root, S: stem, L: young leaves, F: flowers; SG(G) and (P), LG(G) and (P), Y(G) and (P) and Red(G) and (P) are for grain and for pericarp, respectively, from small green, large green, yellow and red cherries, and SG(G), LG(G), Y(G) and R(G) for pericarp from small green, large green, yellow and red cherries respectively. Five micrograms of total RNA was loaded in each lane. Panel A demonstrates the ethidium bromide staining of large ribosomal RNA prior to blotting as a loading control and panel B is an autoradiogram showing the appearance of the CcAP2 mRNA in the specific tissues tested.

Detailed Description of the Invention

As used herein, a "polynucleotide" is a nucleotide sequence such as a nucleic acid fragment. A polynucleotide may be a polymer of RNA or DNA that is single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may comprise one or more segments of cDNA, genomic DNA, synthetic DNA or mixtures thereof.

Similar nucleic acid fragments are characterised, in the present invention, by the percent identity of the amino acid sequences that they encode, to the amino acid sequences disclosed herein, as determined by algorithms commonly used by those skilled in the art. Suitable nucleic acid fragments (or isolated polynucleotides of the first to third aspects of the present invention) encode polypeptides that are at least 70% identical, preferably at least 80% identical, to the amino acid sequences disclosed herein. Preferred nucleic acid fragments encode amino acid sequences that are at least 85% identical to the amino acid sequences disclosed herein. More preferred nucleic acid fragments encode amino acid sequences that are at least 90% identical to the amino acid sequences disclosed herein. Still more preferred are nucleic acid fragments that encode amino acid sequences that are at least 95% identical to the amino acid sequences disclosed herein. Multiple alignment of sequences should be performed using the ClustalW method of alignment (Thompson *et al*, 1994, Nucleic Acids Research, Vol 22, p4673-4680; Higgins & Sharp 1989 Cabios. 5:151-153).

As used herein, the term "similar nucleic acid fragments" refers to polynucleotide sequences in which changes in one or more nucleotide bases result in substitution of one or more amino

acids, but which changes either do not affect the function of the polypeptide encoded by the nucleotide sequence or do not affect the ability of nucleic acid fragment to mediate gene expression by gene silencing via, for example, antisense or co-expression technology. The term "similar nucleic acid fragments" also refers to modified polynucleotide sequences, in which one or more nucleotide bases is / are deleted or inserted, provided that the modifications do either do not affect the function of the polypeptide encoded by the nucleotide sequence or do not affect the ability of nucleic acid fragment to mediate gene expression by gene silencing. It will, therefore, be understood that the scope of the present invention extends beyond the polynucleotide and polypeptide sequences specifically disclosed herein.

Similar nucleic acid fragments may be selected by screening nucleic acid fragments in the form of subfragments or modified nucleic acid fragments, for their ability to affect the level of the polypeptide encoded by the unmodified nucleic acid fragments in the plant or plant cell.

The term "operably linked" refers to the association of two or more nucleic acid fragments on a single nucleic acid fragment so that the function of one is affected by the other. "Regulatory sequences" refer to nucleotide sequences located upstream, within, or downstream, of a coding sequence and which influence transcription, RNA processing or stability, or translation of the coding sequence associated therewith. Regulatory sequences may include promoters, translation leader sequences, introns and polyadenylation recognition sequences. When a regulatory sequence in the form of a promoter is operably linked to a coding sequence, the regulatory sequence is capable of affecting the expression of the coding sequence. Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression" refers to the transcription, and stable accumulation, of sense RNA (mRNA) or antisense RNA derived from the nucleic acid fragments of the present invention. Expression may also refer to the translation of mRNA into a polypeptide. Overexpression refers to the production of a gene product in a transgenic cell, that exceeds the level of production in normal, or non-transformed, cells. "Altered levels" refers to the production of gene product(s) in a transgenic cell in amounts or proportions that differ from that of normal, or non-transformed, cells.

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host cell, resulting in genetically stable inheritance. Host cells containing the transformed nucleic acid fragments are referred to herein as "transgenic cells.

Standard recombinant DNA and molecular cloning techniques as used herein are well known in the art and are described more fully in Sambrook *et al* "Molecular Cloning: A Laboratory Manual"; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989, which is incorporated herein by reference.

Examples

The following Examples illustrate the invention without limiting the invention to the same. In the examples, all parts and percentages are by weight and degrees are in Celsius, unless this is otherwise specified.

In the following Examples, these abbreviations have been used:
PCR : Polymerase chain reaction
RACE : Rapid amplification cDNA ends

From the above discussion and the Examples below, those skilled in the art can ascertain the essential features of the present invention, and without departing from the scope thereof can make various changes and modifications thereto, to adapt it to various usages and conditions as desired.

Production of cDNA libraries and screening

Production of Seed Specific RNA
Coffee cherries of the *Robusta* variety Q121 were harvested 30 WAF (weeks after flowering) at the ICCRI, Indonesia. The pericarps of these cherries were then removed and the remaining perisperm/endosperm material was frozen and ground to a powder in liquid nitrogen. The RNA was extracted from the frozen powder material using the method described previously for the RNA extraction of cacao seeds (Guilloteau, M. *et al*, 2003, Oil bodies in *Theobroma cacao* seeds: cloning and characterisation of cDNA encoding the 15.8

and 16.9 kDa oleosins. Plant Science Vol 164, 597-606). Poly A^+ RNA was prepared from approximately 250µg total RNA using the "PolyA Purist™" kit of AMBION (manufactured by Ambion, Inc.) according to their kit instructions.

Production of First Set of Seed cDNA clones

Approximately 50-100ng of this poly A^+ RNA was then employed in the synthesis of the first strand cDNA using "SuperScript™ II RNase H⁻ reverse transcriptase (GIBCOBRL™)and the SMART™ PCR cDNA synthesis kit (Clontech) as follows. A reaction containing 2µl of 30 WAF poly A^+ RNA, 1 µL CDS oligo (SMART™ PCR cDNA kit, Clontech), 1 µL Smart II oligo (SMART™ PCR cDNA kit, Clontech), and 8 µL deionised H_2O. This mixture was heated to 72°C for 5 minutes and then placed on ice. Then the following was added; 1 µL 10 mM dNTPs, 4 µL SuperScriptII™ 1st stand buffer and 2 µL DTT. This mixture was put at 42°C for 2 minutes then 1 µL of SuperScriptII™ RNaseH⁻ reverse transcriptase (200 units/µL GIBCO BRL™) was added and the mixture was incubated in an air circulating incubator at 42°C for a further 50 minutes.

After the reverse transcription reaction, the following PCR reaction was carried out. 98 µL of the Master Mix described in the SMART™ PCR cDNA kit (Clontech) containing Advantage™ 2 polymerase (Advantage™ 2 PCR kit, ClonTech) was set up on ice and then 3 µL of the 1st strand cDNA synthesis reaction described above was added. This 100 µL PCR reaction was then placed in a MJ Research PTC-150 HB apparatus and the following PCR conditions were run: 95°C for 1 minute, then 16 cycles of 95°C for 15 seconds, 65°C for 30 seconds, 68°C for 6 minutes. The amplified DNA was purified using the Strataprep™ PCR Purification Kit (Stratagene) according to the suppliers' instructions. The DNA, which was eluted in 50 µL deionized water, was then "polished" using the Pfu-1 polymerase reagents contained in the PCR-Script™ Amp cloning kit (Stratagene) as follows; 50 µL DNA, 5 µL 10 mM dNTPs, 6.5 µL 10 x Pfu-1 polishing buffer, 5 µL cloned Pfu-1 DNA polymerase (0.5 U/µl). This reaction was then incubated at 72°C for 30 minutes in a PCR apparatus with a heated cover (Perkin Elmer). Using the protocol described in the pPCR-Script™ Amp kit (Stratagene), the polished (blunted) PCR products were ligated into the Srf-1 digested pPCR-Script™ Amp SK(+) vector in the presence of Srf-1 enzyme and the ligation reaction products were transformed into the XL-10 Gold™ Kan ultracompetent E. coli cells. Selection for transformation with plasmids containing inserts was done using LB-Amp plates

and IPTG and Xgal spread on the surface as described in the pPCR-Script™ Amp kit. White colonies were selected and the clones were named Dav1-1 etc.

Production of Second Set of seed cDNA clones with Size Selected cDNA

Seeds highly express a small number of proteins, such as the seed storage proteins (White *et al*, 2000, Plant Physiology, Vol 124, 1582-1594). When cDNA is prepared from such tissue, the very high level of the storage proteins and other seed specific proteins leads to a high level of cDNA "redundancy", that is, the population of cDNA produced contains high proportions of the same cDNA. In order to reduce the redundancy of cDNA made from coffee seed mRNA, and to selectively characterise long and weakly expressed cDNA, a second cDNA cloning strategy was also used. Using the products of the reverse transcriptase reaction described above, the following PCR reactions was set up using the Advantage™ 2 PCR kit (ClonTech): 3 µL of the reverse transcriptase reaction, 5 µL 10 x Advantage™ 2 PCR buffer, 1 µL dNTP's (10 mM each), 2 µL PCR primer (SMART™ PCR cDNA kit, Clontech), 39 µL deionised water, and 1 µL 50 x Advantage™ 2 polymerase mix. This PCR reaction was then placed in a MJ Research PTC-150 HB apparatus and the following PCR conditions were run: 95°C for 1 minute, then 16 cycles of 95°C for 15 seconds, 65°C for 30 seconds, 68°C for 6 minutes. At the end of the PCR, 1 µL 10% SDS was added with gel loading buffer, the sample was heated to 37°C for ten minutes. The sample was then split for loading onto a 0.7% agarose gel without ethidium bromide: 10% was loaded into a small well beside a DNA marker lane and the other 90% was loaded into a neighbouring large, preparation scale well. After the gel was run, the gel section with the size markers, plus the 10% reaction sample, were stained with ethidium bromide. This stained gel section was then used as a template to generate gel slices containing PCR amplified cDNA of different sizes from the cDNA present in the remaining unstained (preparation) part of the gel. Six gel slices were generated having the indicated size range of PCR fragments; A1A (0.8-1kb), A1B (1-1.5 kb), A2 (1.5-2.25 kb), A3 (2.25-3.25), A4 3.25-4 kb), and A5 (4-6.5 kb).

The DNA in each gel slice was eluted from the agarose using the QIAEX II kit from Qiagen following the suppliers instructions (for samples 3A, 4A, and 5A were heated for 10 minutes at 50°C and 1A, 1B, and 2A were heated for 10 minutes at room temperature). The purified double stranded cDNA was then re-amplified further by PCR with a TAQ enzyme mix which makes fragments having a 3' T overhang as follows: 30 µL of the gel isolated double

stranded cDNA, 5 µL 10 x TAQ buffer (supplied with TAQ PLUS precision polymerase mix, Stratagene), 1 µL 40 mM dNTP's (each 10 mM), 2 µL PCR primer (SMART™ PCR cDNA kit, Clontech), 0.5 µL TAQ PLUS precision polymerase mix (Stratagene) and 11.5 µL deionised water. The PCR reaction conditions were as follows: 95°C for 1 minute then 7 cycles 95°C for 15 seconds, 65°C for 1 minute, 72°C for 8 minutes, then 1 cycle at 95°C for 15 seconds, 65°C for 1 minute, 72°C for 10 minutes.

The PCR amplified DNA produced was then ligated into the vector pCR™-TOPO™ and cloned into TOP10 *E. coli* cells using the TOPO™ TA kit (Invitrogen) as described by the supplier. The clones were named by their order of isolation and their position in the sizing gel (for example, A2-1, A2-2, etc.).

Seed cDNA Screening and Preliminary Identification

The first set of white colonies obtained in Dav-1 library were screened by first determining the size of each insert by PCR amplifying the insert using the primers T3 and T7 which flank the cloning site used and examining the PCR amplified fragments on a gel.

Each white colony were resuspended in 200 µl sterile water and 10-30 µl of this was added to 5 µl 10X Taq polymerase buffer (Stratagene), 1 µl 10 mM dNTP mix, 2.5 µl 20 µM T3 primer, 2.5 µl 20 µM T7 primer, 1 µl DMSO, 0.5 µl Taq polymerase (Stratagene), and H_20 up to 50 µl final volume). The PCR reaction program uses was 94°C for 1 min, then 30 cycles of 94°C for 1 min, 55°C for 1.5 min and 3.5 min at 72°C, and a final cycle of 7 min at 72°C. To reduce redundancy, the PCR inserts of similar size were subjected to digestion by the restriction enzyme Hae III. Those PCR fragments with the same Hae III restriction pattern were not studied further. The plasmids of clones with PCR fragments >500 bp and which had unique Hae III restriction patterns were then purified by using the Qiawall 8 ultra plasmid kit (Qiagen) for 5' end dideoxy sequencing using the appropriate T7 or T3 sequencing primers coded in the flanking vector sequences. Because the inserts were not cloned in a directed fashion, it was first necessary to determine the 5' end of each clone by a ScaI digestion of the purified plasmid DNA (the CDS SMART primer contains a Sca 1 site allowing the orientation of the insert to be determined). The DNA sequence data obtained was subsequently blasted against the non-redundant database protein in GENEBANK to obtain a preliminary annotation of each cDNA clone using the program BLASTX™.

Seed cDNA banks have a high level of redundancy. That is, a small number of seed mRNA have an unusually high level of expression, such as those encoding the seed storage proteins, and therefore their cDNA are very abundant in seed cDNA banks (White *et al*, 2000, Plant Physiology, Vol 124, 1582-1594). Therefore, as soon as the main redundant cDNA's were identified in the first round of sequencing the coffee seed cDNA, a pre-screening step was added for the white insert containing colonies prior to the determination of insert size. Four sequences were very highly expressed and the following specific primers sets were made for each of these redundant sequences,

1) 2S protein, contig 8A 5' AGCAACTGCAGCAAGGTGGAG 3' and contig 8B 5' CGATTTGGCACTGCTGTGGTTC 3' (55°C used in PCR, 114 bp fragment),
2) 2S protein contig 15A 5' GCCCGTGCTCCTGAACCA 3' and contig 15B 5' GTATGGTTGCGGTGGCTGAA 3' (55°C used in PCR,256 bp fragment),
3) Oleosin 15.5 contig 30A 5' ACCCCGCTTTTCGTTAT 3' and contig 30B TCTGGCTACATCTTGAGTTCT 3' (55°C used in PCR, 261 bp fragment), and
4) 11S protein contig 37A 5' GTTTCCAGACCGCCATCAG 3' and contig 37B 5' ATATCCATCCTCTTCCAACACC 3' (59°C used in PCR, 261 bp fragment).

The PCR reactions for this prescreen step were run as follows: 10-30 µl of the white colony in sterile H_2O, 5 µl 10X Taq buffer (Stratagene), 1 µl 10 mM dNTP, 2.5 µl of each primer at 20 µM, 1 µl DMSO, 0.5 µl Taq polymerase (Stratagene 10U/µl) and sterile H_2O was added to produce a final reaction total volume of 50 µl. The PCR program was 1 min at 94°C, then 30 cycles of 1 min at 94°C, 1.5 min at specific temperature for each primer pair, 2.5 min at 72°C, followed by 7 min at 72°C.

Full Length cDNA Insert Sequencing and Sequence Analysis

cDNA clones whose partial sequences showed initial homologies to proteinases and proteinase inhibitors were fully sequenced on both strands using the standard dideoxy primer walking strategy. The sequences are shown under SEQ ID Nos. 1, 3, 5 and 7. The full length sequences obtained were again blasted against the GenBank non redundant protein database using BLASTX to reinforce the preliminary annotation

Sequence identities of sequence pairs were calculated using the ClustalW™ program contained in the MegAlign™ module of the Lasergene™ software package (DNASTAR Inc). The default parameters were chosen as follows: (1- MULTIPLE ALIGNMENT PARAMETERS - Gap penalty 15.00, Gap length penalty 6.66, Delay divergent Seqs (%) 30, DNA transition weight 0.5, Protein Weight Matrix-Gonnet Series, DNA Weight Matrix IUB. 2- PAIRWISE ALIGNMENT PARAMETERS-Slow/Accurate (Gap Penalty 15.00, Gap Length Penalty 6.66), Protein Weight Matrix-Gonnet 250, DNA Weight Matrix-IUB) and the sequences used were either the full length nucleotide sequence of each cDNA or the full ORF (open reading frame) of each cDNA.

TABLE 2

Identity values between the nucleic acid and amino acid sequences of CcCP-1, CcCPI-1, and CcAP-2 and related genes found in the non-redundant protein database of GeneBank and those of WO 02/04617.

cDNA Sequences	nucleotide identity (%)	protein identity (%) (ORF)
CcAP1 vrs TcAP1	2.9	13.3
CcAP1 vrs TcAP2	2.4	9.8
CcAP2 vrs TcAP1	55.0	61.5
CcAP2 vrs TcAP2	55.1	61.3
CcCP-1 vs *Arabidopsis thaliana* putative cysteine proteinase (AY070063)	51.8	64.3
CcCP-1 vs Glycine max cysteine endopeptidase (Z32795)	49.1	61.3
CcCP-1 vs *Vicia sativa* cysteine proteinase precursor (Z99172)	49.0	60.9
CcAP2 vs *Lycopersicon esculentum* aspartic proteinase precursor (L46681)	65.9	71.1
CcAP2 vs *Ipomoea batatas*		

putative aspartic proteinase mRNA (AF259982)	71.7	69.6
CcAP2 vs *Nepenthes alata* NaAP4 mRNA for aspartic proteinase 4 (AB045894)	58.4	66.5
CcCPI-1 vs *Malus x domestica* cystatin (AY176584)	38.8	45.5

Northern-Blot Analysis

Freshly harvested roots, young leaves, stem, flowers and fruit at different stages of development (small green fruit (SG), large green fruit (LG), Yellow fruit (Y) and red fruit (R)) were harvested from *Coffea arabica* var. Catura T-2308 grown under greenhouse conditions (25°C, 70 RH) in Tours, France, and from *Coffea canephora* var. BP409 grown either in Equador or ICCRI, Indonesia. The fresh tissues were frozen immediately in liquid nitrogen and total RNA was isolated from each tissue using the extraction procedure described above. A total of 5 µg of RNA was run on a 1.2% (w/v) denaturing RNA gel containing formaldehyde. The total RNA samples from each plant tissue were heated at 65°C for 15 min in presence of 7 µL "RNA Sample Loading Buffer" (without ethidium bromide, Sigma), and then put immediately on ice for 2 minutes before being loaded onto the 1.2% RNA gel. The gels were run at 60 Volts for 5 hours. The gel was then soaked twice in 10× SSC for 20 min. The RNA in the gel was transferred overnight by capillary transfer to a "Positive TM Membrane" (Qbiogene) in 10× SSC and the RNA was fixed by heating the blot for 30min at 80°C. Probes were generated using "Rediprime™ II random prime labelling system" kit (Amersham) in the presence of (P^{32}) dCTP. Hybridisation was carried out at 65°C for 24 h in hybridisation solution (5X SSC, 40µg/ml Denatured Salmon Sperm DNA, 5% [w/v] SDS, and 5× Denhardt's solution). Then, the membrane was washed twice at 65°C using 2X SSC, 0.1% SDS [w/v] and 1X SSC, 0.1% SDS [w/v] during 30 minutes each.

5' RACE PCR

The cDNA insert of clone A5-812 was found to contain introns. Therefore, to confirm the coding sequence of this protein, it was necessary to isolate a new cDNA containing the complete coding sequence. This was accomplished by using the SMART™ RACE cDNA amplification Kit (Clontech). The first strand cDNA used for the 5' RACE was made as already described for the cDNA libraries above. A gene specific primer rAP2 (5' CATATAATATTAAAAGCACCACCCATAA 3') was designed - this sequence is situated 92 pb from the poly (A) tail of A5-812 clone. This specific primer was then used with the Universal Primer Mix (UPM) in the CLONTECH kit in a PCR reaction under the following conditions; 2.5 µl of first strand cDNA product, 5 µl of 10X Advantage 2 PCR Buffer (CLONTECH), 1 µl of dNTP Mix (10 mM), 1 µl of 50X Advantage 2 Polymerase Mix (CLONTECH), 5 µl of "Universal Primer A Mix" (10X) (CLONTECH), 1 µl of rAP2 (10 µM) and sterile water was added to a final volume of 50 µl. PCR cycling conditions were 20 cycles of 30 sec at 94°C, 30 sec at 68°C and 3 min at 72°C, followed by a final extension reaction for 5 min at 72°C. A fragment of about 1700 pb was obtained, excised from the gel using "CONCERT™ Rapid Gel Extraction kit" (GibcoBRL). The isolated fragment was cloned in the pCR 4-TOPO vector and transformed into Escherichia coli using the Topo-TA cloning kit (Invitrogen). The plasmid obtained was then purified using a plasmid extraction kit (QIAfilter Plasmid Midi Kit, Qiagen, France) and the insert of this plasmid was double strand sequenced.

The DNA of clone A5-442 (AP1) was found to lack the 5' region of the cDNA. To isolate this region a 5' RACE was performed using the SMART™ RACE cDNA amplification Kit (Clontech). A sequence specific primer rAP1 (5'-TGGAGTCACAAGATGTCTCGACGAACTG-3') situated at 396 pb from the poly (A) tail was designed. This specific primer was then used with the Universal Primer Mix (UPM) in the CLONTECH kit in a PCR reaction under the following conditions; 2.5 µl of first strand cDNA, 5 µl of 10X Advantage 2 PCR Buffer (CLONTECH), 1 µl of dNTP Mix (10 mM), 1 µl of 50X Advantage 2 Polymerase Mix (CLONTECH), 5 µl of "Universal Primer A Mix" (10X) (CLONTECH), 1 µl of rAP1, and sterile water was added to a final volume of 50 µl. PCR cycling conditions were 20 cycles of 30 sec at 94°C, 30 sec at 68°C and 3 min at 72°C, followed by a final extension reaction for 5 min at 72°C. A fragment of about 2,000 bp was obtained, excised from the gel using "CONCERT™ Rapid Gel Extraction kit" (GibcoBRL). The isolated fragment was cloned in the pCR 4-TOPO vector and transformed into

Escherichia coli using the Topo-TA cloning kit (Invitrogen). The plasmid obtained was then purified using a plasmid extraction kit (QIAfilter Plasmid Midi Kit, Qiagen, France) and the insert of this plasmid was double strand sequenced.

The Northern blot analysis shown in Figure 1 demonstrates that the coffee cysteine proteinase gene CcCP-1 gene is expressed in the *C. arabica* coffee cherry at all the stages tested, with yellow cherries exhibiting slightly higher levels of expression than the other stages. No expression was detected for this gene in the root, stem or leaves of *C. arabica.*

The Northern blot analysis shown in Figure 2 demonstrates that the coffee cysteine proteinase inhibitor gene CcCPI-1 gene is expressed in the *C. arabica* coffee cherry at all stages tested. However, in contrast to the expression seen for the cysteine proteinase CcCP-1, CcCPI-1 exhibits higher expression in the two early stages of coffee cherry development (small green and large green), and this gene is expressed at lower levels in the two later stages of cherry development. This expression pattern is consistent with the present hypothesis that the cysteine proteinase inhibitor protein (CcCPI-1) controls the activity level of the cysteine proteinase CcCP-1 in the coffee cherry. A controlling protein such as the cysteine proteinase inhibitor protein can be expected to be expressed earlier than its target protein if it is necessary to control the level of activity of its target protein continuously from the time that the target protein is expressed. No expression was detected for this gene in the root, stem or leaves of *C. arabica.* It is noted that the similarity of the expression patterns for CcCP-1 and CcCPI-1 are consistent with the present hypothesis that these proteins interact functionally.

The Northern blot analysis shown in Figure 3 demonstrates that the coffee cysteine proteinase inhibitor gene CcCPI-1 gene is expressed differently in the cherries of *C. canephora* (robusta) versus the cherries of *C. arabica.* First, the data of Figure 3 shows that the CcCPI-1 gene is expressed slightly earlier in *C. arabica.* Secondly, and more importantly, the CcCPI-1 gene is expressed in significantly higher levels in the *C.canephora* cherries. This difference in expression probably affects the level of the cysteine proteinase activity found in *C. arabica* versus *C. canephora* cherries. Because this class of protein is widely associated with insect resistance in plants, it is also likely that the high expression of the CcCPI-1 gene in C. canephora contributes to the higher disease resistance often seen for robusta varieties versus arabica varieties.

The Northern blot analysis shown in Figure 4 demonstrates that the coffee aspartic proteinase CcAP-2 gene is expressed in both the grain and the pericarp of the *C. arabica* coffee cherry at all cherry development stages tested. The CcAP-2 gene also has a relatively high expression in roots. When the film is exposed longer, CcAP-2 expression was also detected in the tissues of *C. arabica* stems, leaf, and flowers.

Overexpression and under-expression of the CcCP-1, CcAP-1 and CcAP-2 proteinase gene sequences and the CcCPI-1 proteinase inhibitor in coffee seeds.

It is expected that the major storage protein profile and the amino acid/peptide profile can be changed in the mature coffee grain by altering, either up or down, the expression of one or more of the genes disclosed herein.

Methods for the overexpression of a gene of interest are well known in the art. Such methods consist of creating a chimeric gene of three major components, 1) a promoter sequence at the 5' end of the gene, preferably in the current application a seed specific promoter such the coffee seed specific promoter described in Marrachini et al. 1999 (Marraccini *et al* 1999 Molecular cloning of the complete 11S seed storage protein gene of *Coffea arabica* and promoter analysis in transgenic tobacco plants, Plant Physiol. Biochem. Vol 37, 273-282, and WO 99/02688), 2) the entire coding sequence of the gene to be expressed, and 3) a 3' control region such as the 3' region from the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. Then, the chimeric gene can be cloned into an *Agrobacterium tumefaciens* transformation vector, and this vector can be transformed into an *Agrobacterium tumefaciens* strain for use in coffee transformation as described by Leroy *et al* 2000, (Leroy *et al* 2000 Genetically modified coffee plants expressing the *Bacillus thuringiensis cry1Ac* gene for resistance to leaf minor. Plant Cell Reports 2000, 19, 382-389). Plants with stable transformation inserts can then be screened for those which overexpress the specific genes used in the transformation experiment specifically in mature seeds using methods such as detection of gene overexpression or protein activity overexpression versus seeds from mock transformed plants.

It is well known in the art that the expression of known gene sequences can be reduced or completely blocked by antisense suppression and of gene expression using nucleic acid fragments representing less than the entire coding region of a gene, and by nucleic acids that do not share 100% sequence identity with the gene to be suppressed. In this case, the

sequences chosen for the particular antisense suppression or cosuppression experiment will replace the full length gene in the chimeric gene construction scheme presented above. The resulting antisense suppression or cosuppression chimeric constructions are again cloned into an *Agrobacterium tumefaciens* transformation vector, and transformed into *Agrobacterium tumefaciens* strain for use in coffee transformation as described above. Plants with stable transformation inserts can then be screened for those with reduced expression of the specific gene sequences used in the seeds of the transformed plants. The reduced expression can be detected by techniques such as northern blotting ;semi quantitative RT-PCR, and/or quantitative RT-PCR.

Another method for reducing, or eliminating, the expression of a gene in plants is to use the small portions of the gene sequences disclosed herein to produce RNA silencing via using RNAi (Hannon, G.J., 2002, Nature, Vol 418, 244-251; Tang *et al*, 2003, Genes Dev, Vol 17, 49-63). In this approach, small regions of one or more of the sequences disclosed herein are cloned into an *Agrobacterium tumefaciens* transformation vector as described above which has a seed specific promoter and an appropriate 3' regulatory region. This new inserted sequence for RNAi should be constructed so that the RNA produced forms an RNA structure *in vivo* which result in the production of small double stranded RNA in the transformed cells and whereby these small double stranded RNA sequences trigger the degradation of the homologous mRNA in these transformed cells.

Screening for naturally occurring variations in the CcCP-1, CcAP-1, CcAP-2, CcCPI-1 genes and creating new mutations in these genes.

The sequences disclosed herein can be used to screen natural populations for allelic variants in these genes. This can be accomplished by using the CcCP-1, CcAP-2, CcCPI-1 sequences as probes in a search for naturally occurring RFLP's (restriction fragment length polymorphisms) in genomic DNA from different coffee plant varieties. A more powerful method to find allelic variants is to use the mutation screening technology associated with the TILLING method (Till, B.J., *et al* 2003 Large scale discovery of induced point mutations with high-thruput TILLING. Genome Research Vol 13, 524-530). In this case, once a specific gene sequence has been isolated and cloned, such as CcCP-1, CcAP-2, CcCPI-1 sequences herein, the mutation screening technique associated with the TILLING method can be used to identify sequence variants between the cloned sequence and the corresponding cDNA or genomic sequence in different varieties. Using PCR primer pairs coding for DNA

segments of 700-1100 base pairs, the known cloned gene can be scanned for naturally occurring sequence variations in different varieties. In the ideal situation, one or more sequence variants could also be correlated with a particular phenotypic variation thereby identifying a genetic marker for this phenotypic variant.

Additionally, using the sequences disclosed herein for CcCP-1, CcAP-2 and CcCPI-1, application of the full TILLING method can be used to create and detect new mutants in these genes and thus produce plants containing these specific mutants. For example, using the full TILLING method, coffee plants could be created which have specific mutations, such as a missense mutation in the coding sequence which inactivates the gene target of interest.

CLAIMS:

1. An isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide having cysteine proteinase activity, wherein the amino acid sequence of the polypeptide and the amino acid sequence of SEQ ID No. 2 have at least 70%, preferably at least 80%, sequence identity based on the ClustalW alignment method; or the complement of the nucleotide sequence, wherein the complement contains the same number of nucleotides as the nucleotide sequence, and the complement and the nucleotide sequence are 100% complementary.

2. The polynucleotide of Claim 1, wherein the amino acid sequence of the polypeptide and the amino acid sequence of SEQ ID No. 2 have at least 85%, preferably at least 90%, optionally at least 95%, sequence identity based on the ClustalW alignment method.

3. The polynucleotide of Claim 1, wherein the nucleotide sequence comprises the nucleotide sequence of SEQ ID No. 1.

4. The polynucleotide of Claim 1, wherein the polypeptide comprises the amino acid sequence of SEQ ID No. 2.

5. An isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide having cysteine proteinase inhibitor activity, wherein the amino acid sequence of the polypeptide and the amino acid sequence of SEQ ID No. 4 have at least 70%, preferably at least 80%, sequence identity based on the ClustalW alignment method; or the complement of the nucleotide sequence, wherein the complement contains the same number of nucleotides as the nucleotide sequence, and the complement and the nucleotide sequence are 100% complementary.

6. The polynucleotide of Claim 5, wherein the amino acid sequence of the polypeptide and the amino acid sequence of SEQ ID No. 4 have at least 85%, preferably at least 90%, optionally at least 95%, sequence identity based on the ClustalW alignment method.

7. The polynucleotide of Claim 5, wherein the nucleotide sequence comprises the nucleotide sequence of SEQ ID No. 3.

8. The polynucleotide of Claim 5, wherein the polypeptide comprises the amino acid sequence of SEQ ID No. 4.

9. An isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide having aspartic endoproteinase activity, wherein the amino acid sequence of the polypeptide and the amino acid sequence selected from SEQ ID No. 6 or 8, preferably SEQ ID No. 8, have at least 75%, preferably at least 80%, sequence identity based on the ClustalW alignment method, or the complement of the nucleotide sequence, wherein the complement contains the same number of nucleotides as the nucleotide sequence, and the complement and the nucleotide sequence are 100% complementary.

10. The polynucleotide of Claim 9, wherein the amino acid sequence of the polypeptide and the amino acid sequence selected from SEQ ID No. 6 or 8, preferably SEQ ID No. 8, have at least 85%, preferably at least 90%, optionally at least 95%, sequence identity based on the ClustalW alignment method.

11. The polynucleotide of Claim 9, wherein the nucleotide sequence comprises the nucleotide sequence of SEQ ID No. 5 or 7, preferably SEQ ID No. 7.

12. The polynucleotide of Claim 9, wherein the polypeptide comprises the amino acid sequence of SEQ ID No. 6 or 8, preferably SEQ ID No.8.

13. A vector comprising the polynucleotide of any one of Claims 1 to 12.

14. A recombinant DNA construct comprising the polynucleotide of any one of Claims 1 to 12 operably linked to a regulatory sequence.

15. A method for transforming a cell comprising transforming a cell with the polynucleotide of any one of Claims 1 to 12.

16. A cell comprising the recombinant DNA construct of Claim 14.

17. The cell of Claim 16, which is a prokaryotic cell, an eukaryotic cell or a plant cell, preferably a coffee cell.

18. A transgenic plant comprising the cell of Claim 16 or 17.

19. A method for modulating coffee flavour precursor levels in green coffee grains, the method comprising introducing into the coffee plant the recombinant DNA construct of Claim 14.

Abstract

Modulation of Coffee Flavour Precursor Levels in Green Coffee Grains

The present invention relates to isolated polynucleotides encoding cysteine proteinases; cysteine proteinase inhibitors; and aspartic endoproteinases.

The invention also relates to a transformed host cell, preferably a plant cell, in which over- or under- expression of these polynucleotides result in altered levels of coffee flavour precursor levels, specifically, amino group-containing molecules such as amino acids, peptides and proteins, in green coffee grains.

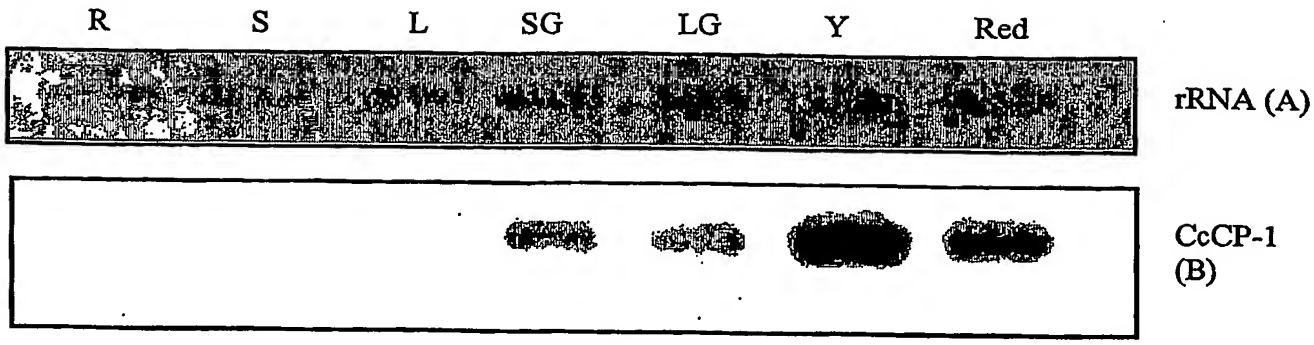

Figure 1: Northern blot analysis of the expression of the cysteine proteinase (CcCP1) gene in different tissues of *Coffea arabica*.

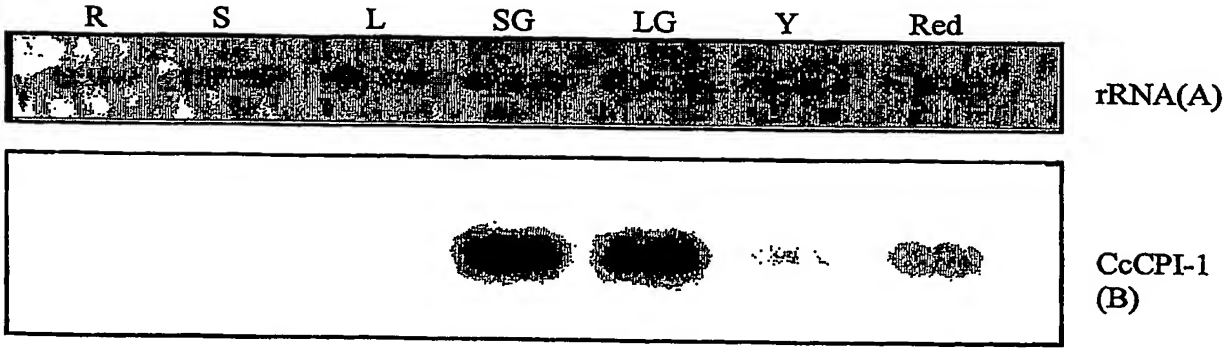

Figure 2: Northern blot analysis of the expression of the cysteine proteinase inhibitor (CcCPI-1) gene in different tissues of *Coffea arabica*.

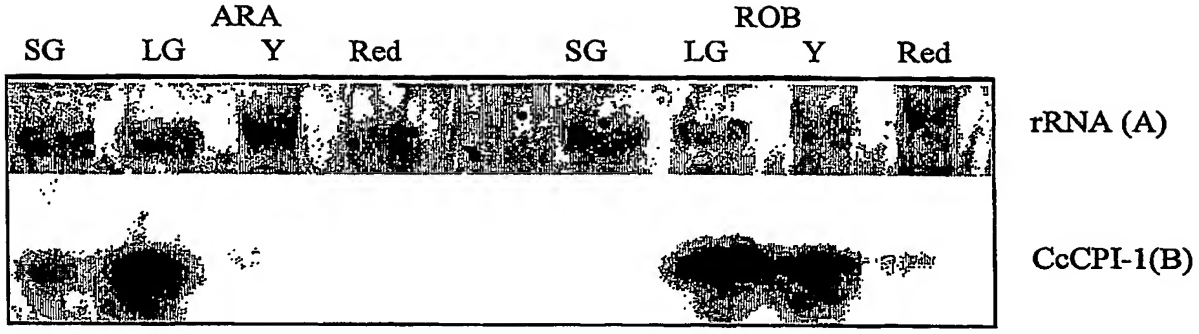

Figure 3: Northern blot analysis of the expression of the cysteine proteinase inhibitor gene (CcCPI-1) at different cherry development stages for *Coffea arabica* (ARA) and *Coffea canephora* (ROB).

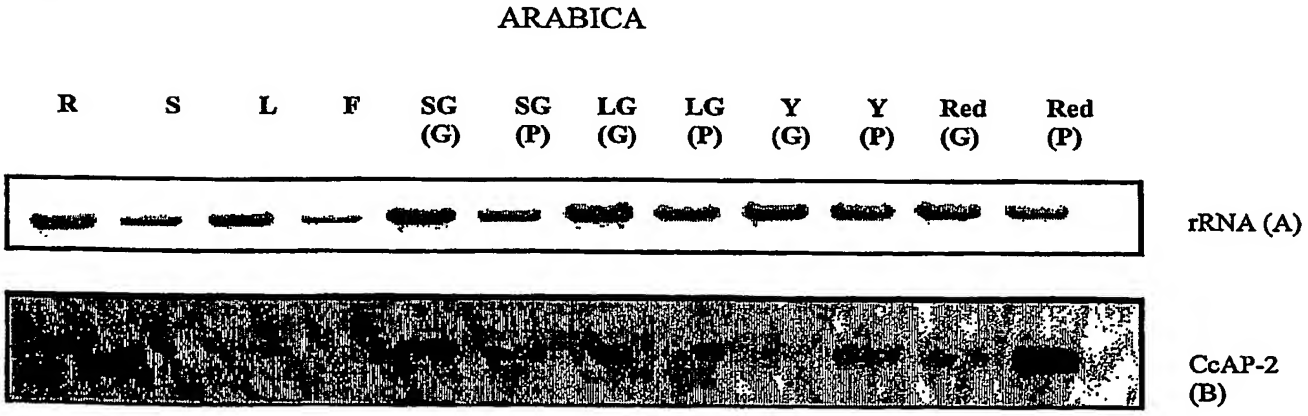

Figure 4: Northern blot analysis of the expression of the aspartic proteinase 2 (CcAP2) gene in different tissues of *Coffea arabica*.

SEQUENCE LISTING

<110> Tours Nestlé Research Center

<120> Implication Of Proteinase And Proteinase Inhibitor In Coffee Flavour

<130> Patent Proteinase and Proteinase Inhibitor Coffee

<160> 8

<170> PatentIn version 3.1

<210> 1
<211> 1543
<212> DNA
<213> Coffea canephora

<220>
<221> mRNA
<222> (1)..(1543)
<223>

<220>
<221> CDS
<222> (122)..(1315)
<223>

<400> 1
aagcagtggt aacaacgcag agtacgcggg ggacactcct ccccgttcca ttccagacca 60

```
gggtccaaaa ccaccgtcca agagaggagc agactgcaga gtgatacata caggcacaaa      120

g atg atg atg aca agc gga ggt ctg atg cta acc tgc act ctg gct att      169
  Met Met Met Thr Ser Gly Gly Leu Met Leu Thr Cys Thr Leu Ala Ile
  1               5                  10                  15

acc ctc tta tcc tgc gca ctc atc tct tca acc act ttc caa cat gaa      217
Thr Leu Leu Ser Cys Ala Leu Ile Ser Ser Thr Thr Phe Gln His Glu
            20                  25                  30

att cag tat cga gta caa gac ccg tta atg ata cgc caa gtc acc gac      265
Ile Gln Tyr Arg Val Gln Asp Pro Leu Met Ile Arg Gln Val Thr Asp
        35                  40                  45

aat cac cac cac cgc cac cac cca ggt agg tct tct gca aac cat cgt      313
Asn His His His Arg His His Pro Gly Arg Ser Ser Ala Asn His Arg
    50                  55                  60

cta ctg ggc acc acc aca gag gtt cac ttc aag tcc ttc gtg gag gag      361
Leu Leu Gly Thr Thr Thr Glu Val His Phe Lys Ser Phe Val Glu Glu
65                  70                  75                  80

tac gag aaa act tac tct acg cac gag gag tac gtg cac cgc ctg ggg      409
Tyr Glu Lys Thr Tyr Ser Thr His Glu Glu Tyr Val His Arg Leu Gly
                85                  90                  95

att ttc gcc aag aac ctc atc aag gcc gcg gag cac cag gcc atg gac      457
Ile Phe Ala Lys Asn Leu Ile Lys Ala Ala Glu His Gln Ala Met Asp
            100                 105                 110

ccc tcc gca atc cac ggc gtc acc cag ttc tct gat ctc acc gag gag      505
Pro Ser Ala Ile His Gly Val Thr Gln Phe Ser Asp Leu Thr Glu Glu
        115                 120                 125

gag ttt gag gct acg tac atg ggc ctt aaa ggt ggc gct gga gtt ggt      553
Glu Phe Glu Ala Thr Tyr Met Gly Leu Lys Gly Gly Ala Gly Val Gly
    130                 135                 140

ggg acc acc cag ctg ggg aaa gat gat ggg gat gag agt gca gca gag      601
Gly Thr Thr Gln Leu Gly Lys Asp Asp Gly Asp Glu Ser Ala Ala Glu
145                 150                 155                 160

gtg atg atg gat gta tct gat ttg ccg gag agt ttt gat tgg aga gaa      649
Val Met Met Asp Val Ser Asp Leu Pro Glu Ser Phe Asp Trp Arg Glu
                165                 170                 175

aaa ggt gct gtg acc gaa gtg aag acg cag gga aga tgt gga tcg tgt      697
Lys Gly Ala Val Thr Glu Val Lys Thr Gln Gly Arg Cys Gly Ser Cys
            180                 185                 190

tgg gct ttt agt aca act gga gct att gaa gga gct aat ttc att gca      745
Trp Ala Phe Ser Thr Thr Gly Ala Ile Glu Gly Ala Asn Phe Ile Ala
        195                 200                 205

act ggc aag ctt ctc agc cta agt gaa cag cag ctt gtg gat tgt gat      793
Thr Gly Lys Leu Leu Ser Leu Ser Glu Gln Gln Leu Val Asp Cys Asp
    210                 215                 220

cat atg tgt gat tta aaa gaa aaa gat gac tgt gat gat gga tgc tcc      841
His Met Cys Asp Leu Lys Glu Lys Asp Asp Cys Asp Asp Gly Cys Ser
225                 230                 235                 240

gga ggg cta atg aca act gct ttc aac tac ttg ata gag gca gga ggt      889
Gly Gly Leu Met Thr Thr Ala Phe Asn Tyr Leu Ile Glu Ala Gly Gly
                245                 250                 255

ata gag gag gag gta acc tat ccc tac act ggg aaa cgc gga gaa tgc      937
Ile Glu Glu Glu Val Thr Tyr Pro Tyr Thr Gly Lys Arg Gly Glu Cys
            260                 265                 270
```

```
aaa ttc aat cct gag aaa gtt gcg gtg aaa gtg cgg aat ttc gca aaa      985
Lys Phe Asn Pro Glu Lys Val Ala Val Lys Val Arg Asn Phe Ala Lys
        275                 280                 285

atc cct gag gat gag agt caa att gct gcc aat gta gtg cat aat ggc     1033
Ile Pro Glu Asp Glu Ser Gln Ile Ala Ala Asn Val Val His Asn Gly
    290                 295                 300

ccg ctt gct att gga ttg aat gcg gta ttc atg caa act tac atc ggg     1081
Pro Leu Ala Ile Gly Leu Asn Ala Val Phe Met Gln Thr Tyr Ile Gly
305                 310                 315                 320

ggt gtg tca tgt cct ctt att tgt gac aaa aag agg atc aac cat ggt     1129
Gly Val Ser Cys Pro Leu Ile Cys Asp Lys Lys Arg Ile Asn His Gly
                325                 330                 335

gtt ctt ctt gtg ggc tat ggt tct aga ggc ttc tca atc ctt agg ctt     1177
Val Leu Leu Val Gly Tyr Gly Ser Arg Gly Phe Ser Ile Leu Arg Leu
            340                 345                 350

ggc tac aag cca tac tgg att atc aag aac tca tgg ggg aag cgt tgg     1225
Gly Tyr Lys Pro Tyr Trp Ile Ile Lys Asn Ser Trp Gly Lys Arg Trp
        355                 360                 365

ggc gaa cat ggt tgc tac cgg ctt tgt cga ggg cac aac atg tgt gga     1273
Gly Glu His Gly Cys Tyr Arg Leu Cys Arg Gly His Asn Met Cys Gly
    370                 375                 380

atg agc aca atg gtt tca gct gtg gtg aca cag acc tct tga             1315
Met Ser Thr Met Val Ser Ala Val Val Thr Gln Thr Ser
385                 390                 395

taccaaaaca tctctgctct tcagaggttg tatacaaggt ggtttgctct tggaagatct   1375

tatcatgttt tcgaaatatt taggtttgta taatatgaag ggtagagagt aataagaacc   1435

aaacaaaagt tcaggcctgt ttctgatagg aatggaatat gatcggagtc atttgttact   1495

ggatcacaaa aaaaaatcca aaaaaaaaaa aaaaaaaaaa aaaaaaaa               1543
```

<210> 2

<211> 397

<212> PRT

<213> Coffea canephora

<400> 2

```
Met Met Met Thr Ser Gly Gly Leu Met Leu Thr Cys Thr Leu Ala Ile
1               5                   10                  15

Thr Leu Leu Ser Cys Ala Leu Ile Ser Ser Thr Thr Phe Gln His Glu
            20                  25                  30

Ile Gln Tyr Arg Val Gln Asp Pro Leu Met Ile Arg Gln Val Thr Asp
        35                  40                  45

Asn His His His Arg His His Pro Gly Arg Ser Ser Ala Asn His Arg
    50                  55                  60
```

Leu Leu Gly Thr Thr Thr Glu Val His Phe Lys Ser Phe Val Glu Glu
65 70 75 80

Tyr Glu Lys Thr Tyr Ser Thr His Glu Glu Tyr Val His Arg Leu Gly
85 90 95

Ile Phe Ala Lys Asn Leu Ile Lys Ala Ala Glu His Gln Ala Met Asp
100 105 110

Pro Ser Ala Ile His Gly Val Thr Gln Phe Ser Asp Leu Thr Glu Glu
115 120 125

Glu Phe Glu Ala Thr Tyr Met Gly Leu Lys Gly Gly Ala Gly Val Gly
130 135 140

Gly Thr Thr Gln Leu Gly Lys Asp Asp Gly Asp Glu Ser Ala Ala Glu
145 150 155 160

Val Met Met Asp Val Ser Asp Leu Pro Glu Ser Phe Asp Trp Arg Glu
165 170 175

Lys Gly Ala Val Thr Glu Val Lys Thr Gln Gly Arg Cys Gly Ser Cys
180 185 190

Trp Ala Phe Ser Thr Thr Gly Ala Ile Glu Gly Ala Asn Phe Ile Ala
195 200 205

Thr Gly Lys Leu Leu Ser Leu Ser Glu Gln Gln Leu Val Asp Cys Asp
210 215 220

His Met Cys Asp Leu Lys Glu Lys Asp Asp Cys Asp Asp Gly Cys Ser
225 230 235 240

Gly Gly Leu Met Thr Thr Ala Phe Asn Tyr Leu Ile Glu Ala Gly Gly
245 250 255

Ile Glu Glu Glu Val Thr Tyr Pro Tyr Thr Gly Lys Arg Gly Glu Cys
260 265 270

Lys Phe Asn Pro Glu Lys Val Ala Val Lys Val Arg Asn Phe Ala Lys
275 280 285

Ile Pro Glu Asp Glu Ser Gln Ile Ala Ala Asn Val Val His Asn Gly
290 295 300

Pro Leu Ala Ile Gly Leu Asn Ala Val Phe Met Gln Thr Tyr Ile Gly
305 310 315 320

Gly Val Ser Cys Pro Leu Ile Cys Asp Lys Lys Arg Ile Asn His Gly
325 330 335

Val Leu Leu Val Gly Tyr Gly Ser Arg Gly Phe Ser Ile Leu Arg Leu
340 345 350

Gly Tyr Lys Pro Tyr Trp Ile Ile Lys Asn Ser Trp Gly Lys Arg Trp
355 360 365

Gly Glu His Gly Cys Tyr Arg Leu Cys Arg Gly His Asn Met Cys Gly
370 375 380

Met Ser Thr Met Val Ser Ala Val Val Thr Gln Thr Ser
385 390 395

<210> 3

<211> 726

<212> DNA

<213> Coffea canephora

<220>

<221> mRNA

<222> (1)..(726)

<223>

<220>

<221> CDS

<222> (79)..(498)

<223>

<400> 3

ggcgcaacaa acattgaaag aaaatcaaga acccaaaaaa accccacaag aaaaaaagaa 60

aaagaagaag aaaagcca atg gca aaa cca tcg tca tct cta ctc aca ctt 111
Met Ala Lys Pro Ser Ser Ser Leu Leu Thr Leu
1 5 10

cct tcc ttt ctt ctg atc ttt ttc att ctt gca cta ttt tcc acc acc 159
Pro Ser Phe Leu Leu Ile Phe Phe Ile Leu Ala Leu Phe Ser Thr Thr
15 20 25

ctc caa gtt aat gcc ttg gga agg aaa gtg gga gca agg gag aag att 207
Leu Gln Val Asn Ala Leu Gly Arg Lys Val Gly Ala Arg Glu Lys Ile
30 35 40

gag gat gtg aag agc aac aaa gaa gtt caa gaa ctt ggg gaa tat tgt 255
Glu Asp Val Lys Ser Asn Lys Glu Val Gln Glu Leu Gly Glu Tyr Cys
45 50 55

gtt tct gag tac aac aag agt ttg cgg aag aag aac aac gaa agt ggt 303
Val Ser Glu Tyr Asn Lys Ser Leu Arg Lys Lys Asn Asn Glu Ser Gly
60 65 70 75

gct cct ata atc ttc aca tct gtg gtg gag gct gag aag cag gtg gtt 351
Ala Pro Ile Ile Phe Thr Ser Val Val Glu Ala Glu Lys Gln Val Val
80 85 90

gct ggg atc aaa tat tat ctc aag att aag gcc acc act tct tct ggg 399
Ala Gly Ile Lys Tyr Tyr Leu Lys Ile Lys Ala Thr Thr Ser Ser Gly
95 100 105

gtt ccc aag gtt tac gat gcc att gtg gtg gtt cgg cct tgg gtt cat 447
Val Pro Lys Val Tyr Asp Ala Ile Val Val Val Arg Pro Trp Val His
110 115 120

act aag cca agg cag ttg ctc aac ttc tcc cct tcc cct gcc act aaa 495
Thr Lys Pro Arg Gln Leu Leu Asn Phe Ser Pro Ser Pro Ala Thr Lys
125 130 135

tga agaagaaaat gttgaaaaag ttggaactgt ttgggagatc taatctgatg 548

attattagta cctttcagtg caaattctct ttgctgttaa gtgttcggtt ttttttttt 608

ccctgtgtct atttatgacc gtggtcatga tgatatggtg tatgatccag taataattaa 668

aatctgttgc ataaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaa 726

<210> 4

<211> 139

<212> PRT

<213> Coffea canephora

<400> 4

Met Ala Lys Pro Ser Ser Ser Leu Leu Thr Leu Pro Ser Phe Leu Leu
1 5 10 15

Ile Phe Phe Ile Leu Ala Leu Phe Ser Thr Thr Leu Gln Val Asn Ala
20 25 30

Leu Gly Arg Lys Val Gly Ala Arg Glu Lys Ile Glu Asp Val Lys Ser
35 40 45

Asn Lys Glu Val Gln Glu Leu Gly Glu Tyr Cys Val Ser Glu Tyr Asn
50 55 60

Lys Ser Leu Arg Lys Lys Asn Asn Glu Ser Gly Ala Pro Ile Ile Phe
65 70 75 80

Thr Ser Val Val Glu Ala Glu Lys Gln Val Val Ala Gly Ile Lys Tyr
85 90 95

Tyr Leu Lys Ile Lys Ala Thr Thr Ser Ser Gly Val Pro Lys Val Tyr
100 105 110

Asp Ala Ile Val Val Val Arg Pro Trp Val His Thr Lys Pro Arg Gln
115 120 125

Leu Leu Asn Phe Ser Pro Ser Pro Ala Thr Lys
130 135

<210> 5

<211> 2282

<212> DNA

<213> Coffea canephora

<220>

<221> mRNA

<222> (1)..(2282)

<223>

<220>

<221> CDS

<222> (439)..(1731)

<223>

<400> 5

actcactata ctttgcattc tcttcaccat tctccctcaa aactccctcc aacattcttt 60

tccttggttt tttcatctat ccctcctata aaaatcgatt attttgttct tttacctctt 120

aaaaatccat tcttggaatt catttatcca tatacaccat acttgtgcat gtccctttg 180

gttgttttgc ttttgtgata agtaattgtt ggtttattgg tttttcatga tggctccgga 240

tctaagaaga aatgggtcgg tagtagcttt agccctgtta gtctctctgg ttgttaatgg 300

tgttattttt gatgtagaag gtaacaataa tgtggttttt gaggtggaac ataaatttaa 360

agggagaagg aatgagaatg gaggaagagg gtctttttga cttcactcaa ggctcatgat 420

tcccaccgcc atggcaga atg ctt gca gcc ctt gac atg cct ttg ggt ggc 471
Met Leu Ala Ala Leu Asp Met Pro Leu Gly Gly
1 5 10

aat ggt tcc cct aca gat gca gcg ctc tat ttc act aag ctt tcg att 519
Asn Gly Ser Pro Thr Asp Ala Ala Leu Tyr Phe Thr Lys Leu Ser Ile
15 20 25

ggg act cct cct cag gat tat tat gtg caa gtg gat aca gga agt gac 567
Gly Thr Pro Pro Gln Asp Tyr Tyr Val Gln Val Asp Thr Gly Ser Asp
30 35 40

att ctc tgg gta aac tgt gct ggt tgt gtc aga tgc ccc aag aaa agc 615
Ile Leu Trp Val Asn Cys Ala Gly Cys Val Arg Cys Pro Lys Lys Ser
45 50 55

agt ctt ggt att gac ttg act cta tat gac atg aaa gcc tcc agc acc 663
Ser Leu Gly Ile Asp Leu Thr Leu Tyr Asp Met Lys Ala Ser Ser Thr
60 65 70 75

```
ggg aga ctt gtt act tgt gat caa gac ttt tgc ttg tct gca ttc aat      711
Gly Arg Leu Val Thr Cys Asp Gln Asp Phe Cys Leu Ser Ala Phe Asn
                80                  85                  90

gcc cca gcc tct gat tgc aag gtt ggt aac ccc tgt gca tat tct gtt      759
Ala Pro Ala Ser Asp Cys Lys Val Gly Asn Pro Cys Ala Tyr Ser Val
            95                  100                 105

act tac gga gac ggg agc tca acc ggc gga tat ttt gtc aga gac tat      807
Thr Tyr Gly Asp Gly Ser Ser Thr Gly Gly Tyr Phe Val Arg Asp Tyr
        110                 115                 120

gca aaa ctt aat caa ctg acg gga aat ctt caa acc ata ccc atg aat      855
Ala Lys Leu Asn Gln Leu Thr Gly Asn Leu Gln Thr Ile Pro Met Asn
    125                 130                 135

ggt agt ata gtg ttt ggg tgt tca tct caa caa tct gga gag cta ggg      903
Gly Ser Ile Val Phe Gly Cys Ser Ser Gln Gln Ser Gly Glu Leu Gly
140                 145                 150                 155

tca tct act gaa gca gtt gat ggc ata att ggt ttt gga caa gca aat      951
Ser Ser Thr Glu Ala Val Asp Gly Ile Ile Gly Phe Gly Gln Ala Asn
                160                 165                 170

tca tct att att tca cag ctt gct tca gca gga aag gtt aaa aaa ata      999
Ser Ser Ile Ile Ser Gln Leu Ala Ser Ala Gly Lys Val Lys Lys Ile
            175                 180                 185

ttt tca cat tgc ttg gat ggt atc aat gga gga ggc ata ttt gct att     1047
Phe Ser His Cys Leu Asp Gly Ile Asn Gly Gly Gly Ile Phe Ala Ile
        190                 195                 200

gga caa gta gtg cag cca aaa cta aag aca aca cca ttg gtc cca aat     1095
Gly Gln Val Val Gln Pro Lys Leu Lys Thr Thr Pro Leu Val Pro Asn
    205                 210                 215

gag gca cat tat aat gtt gtt ctg aac gca att gag gtg ggt ggc gac     1143
Glu Ala His Tyr Asn Val Val Leu Asn Ala Ile Glu Val Gly Gly Asp
220                 225                 230                 235

gtt cta aac ctt ccc tca gat gta tta gga ggt gga tct gga agt ggt     1191
Val Leu Asn Leu Pro Ser Asp Val Leu Gly Gly Gly Ser Gly Ser Gly
                240                 245                 250

aca ata ata gac agt ggt aca acc ttg gct tat ctt cct gat gat gtc     1239
Thr Ile Ile Asp Ser Gly Thr Thr Leu Ala Tyr Leu Pro Asp Asp Val
            255                 260                 265

tat act cca ctt atg gaa aag att acg gca tcc caa tcc aac ttg aaa     1287
Tyr Thr Pro Leu Met Glu Lys Ile Thr Ala Ser Gln Ser Asn Leu Lys
        270                 275                 280

atc cat att gtt gaa aat cag ttc aag tgc ttt gtc tat agt gga aat     1335
Ile His Ile Val Glu Asn Gln Phe Lys Cys Phe Val Tyr Ser Gly Asn
    285                 290                 295

gtt gat gat gga ttt cca gtt gta tct ttc cac ttt gag gat tca ctt     1383
Val Asp Asp Gly Phe Pro Val Val Ser Phe His Phe Glu Asp Ser Leu
300                 305                 310                 315

tct ttg aca gtt tat ccc cat gaa tat ctc ttt gat ctt cat gat gat     1431
Ser Leu Thr Val Tyr Pro His Glu Tyr Leu Phe Asp Leu His Asp Asp
                320                 325                 330

caa tgg tgt att ggt tgg cag aat aag ggt atg cag aca aga gat gga     1479
Gln Trp Cys Ile Gly Trp Gln Asn Lys Gly Met Gln Thr Arg Asp Gly
            335                 340                 345

agg gaa gta act ctt ttg gga gat ctt gta ctc gca aac aag ctt gtt     1527
```

Arg Glu Val Thr Leu Leu Gly Asp Leu Val Leu Ala Asn Lys Leu Val
350 355 360

tcg tat gat ctt gaa aat caa acc att gga tgg gct gaa tat aat tgc 1575
Ser Tyr Asp Leu Glu Asn Gln Thr Ile Gly Trp Ala Glu Tyr Asn Cys
365 370 375

tct tcg agc atc aaa ttg aga gac gag aag tca gga aac gtg tat gct 1623
Ser Ser Ser Ile Lys Leu Arg Asp Glu Lys Ser Gly Asn Val Tyr Ala
380 385 390 395

gtg ggc tct cat atc att tct tca gct cgc ggc ctg aat gct gga aag 1671
Val Gly Ser His Ile Ile Ser Ser Ala Arg Gly Leu Asn Ala Gly Lys
400 405 410

gct cta agg ttc cta ttg tta atc atc aca tca ttg ttg cat gca ctt 1719
Ala Leu Arg Phe Leu Leu Leu Ile Ile Thr Ser Leu Leu His Ala Leu
415 420 425

ttg atc cca tga acatttaaaa tcatactagc tgagaaggag gcattatgat 1771
Leu Ile Pro
430

agcgtaccat ggtactcata gtgatcaggc atcttgctga ttctttggac cattataatt 1831

tctcatgtgt ttaaagtgaa agtcagttcg tcgagacatc ttgtgactcc ataatcttct 1891

tgatcaagct gaactctact cacaaaacca tagctaattc ttttgatctc aaaagagaaa 1951

taggctctgc aaaaggattt cggaggttga tgttgaacat tcttcttatt tggatgttat 2011

tgatacccca gatgattaag gaaagcctat aggaaacaga tggtgggaag gagtatacat 2071

tctttctgac tctttggaac ttcctagcgt atacacatat ttcacacgga atgtatctta 2131

taattcatct gttctttctg tttattgtca acttgtttca aatgattgga gtagctgcaa 2191

taatcaactc ggatggtggt tcatgcttaa ggctcgtctt gcctcattgt taagacgtga 2251

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa a 2282

<210> 6

<211> 430

<212> PRT

<213> Coffea canephora

<400> 6

Met Leu Ala Ala Leu Asp Met Pro Leu Gly Gly Asn Gly Ser Pro Thr
1 5 10 15

Asp Ala Ala Leu Tyr Phe Thr Lys Leu Ser Ile Gly Thr Pro Pro Gln
20 25 30

Asp Tyr Tyr Val Gln Val Asp Thr Gly Ser Asp Ile Leu Trp Val Asn
35 40 45

Cys Ala Gly Cys Val Arg Cys Pro Lys Lys Ser Ser Leu Gly Ile Asp
50 55 60

Leu Thr Leu Tyr Asp Met Lys Ala Ser Ser Thr Gly Arg Leu Val Thr
65 70 75 80

Cys Asp Gln Asp Phe Cys Leu Ser Ala Phe Asn Ala Pro Ala Ser Asp
85 90 95

Cys Lys Val Gly Asn Pro Cys Ala Tyr Ser Val Thr Tyr Gly Asp Gly
100 105 110

Ser Ser Thr Gly Gly Tyr Phe Val Arg Asp Tyr Ala Lys Leu Asn Gln
115 120 125

Leu Thr Gly Asn Leu Gln Thr Ile Pro Met Asn Gly Ser Ile Val Phe
130 135 140

Gly Cys Ser Ser Gln Gln Ser Gly Glu Leu Gly Ser Ser Thr Glu Ala
145 150 155 160

Val Asp Gly Ile Ile Gly Phe Gly Gln Ala Asn Ser Ser Ile Ile Ser
165 170 175

Gln Leu Ala Ser Ala Gly Lys Val Lys Lys Ile Phe Ser His Cys Leu
180 185 190

Asp Gly Ile Asn Gly Gly Gly Ile Phe Ala Ile Gly Gln Val Val Gln
195 200 205

Pro Lys Leu Lys Thr Thr Pro Leu Val Pro Asn Glu Ala His Tyr Asn
210 215 220

Val Val Leu Asn Ala Ile Glu Val Gly Gly Asp Val Leu Asn Leu Pro
225 230 235 240

Ser Asp Val Leu Gly Gly Gly Ser Gly Ser Gly Thr Ile Ile Asp Ser
245 250 255

Gly Thr Thr Leu Ala Tyr Leu Pro Asp Asp Val Tyr Thr Pro Leu Met
260 265 270

Glu Lys Ile Thr Ala Ser Gln Ser Asn Leu Lys Ile His Ile Val Glu
275 280 285

Asn Gln Phe Lys Cys Phe Val Tyr Ser Gly Asn Val Asp Asp Gly Phe
290 295 300

Pro Val Val Ser Phe His Phe Glu Asp Ser Leu Ser Leu Thr Val Tyr
305 310 315 320

Pro His Glu Tyr Leu Phe Asp Leu His Asp Asp Gln Trp Cys Ile Gly
325 330 335

Trp Gln Asn Lys Gly Met Gln Thr Arg Asp Gly Arg Glu Val Thr Leu
340 345 350

Leu Gly Asp Leu Val Leu Ala Asn Lys Leu Val Ser Tyr Asp Leu Glu
355 360 365

Asn Gln Thr Ile Gly Trp Ala Glu Tyr Asn Cys Ser Ser Ser Ile Lys
370 375 380

Leu Arg Asp Glu Lys Ser Gly Asn Val Tyr Ala Val Gly Ser His Ile
385 390 395 400

Ile Ser Ser Ala Arg Gly Leu Asn Ala Gly Lys Ala Leu Arg Phe Leu
405 410 415

Leu Leu Ile Ile Thr Ser Leu Leu His Ala Leu Leu Ile Pro
420 425 430

<210> 7

<211> 1819

<212> DNA

<213> Coffea canephora

<220>

<221> mRNA

<222> (1)..(1732)

<223>

<220>

<221> CDS

<222> (79)..(1602)

<223>

<400> 7

cttactgact ctcgtatatt attcaatcta tcttttgagt tttgcaagag cccatcaagc 60

atcaaggcat aaccaacg atg gag agg agg tac ctt tgg gca gca ttt gta 111
Met Glu Arg Arg Tyr Leu Trp Ala Ala Phe Val
1 5 10

tta ggg gcg att gtg tgt tct cta ttt cct ctt cct tct gaa gga tta 159
Leu Gly Ala Ile Val Cys Ser Leu Phe Pro Leu Pro Ser Glu Gly Leu
15 20 25

aag cga att agc ctg aaa aaa aaa ccc tta gat att caa agc ata aga 207
Lys Arg Ile Ser Leu Lys Lys Lys Pro Leu Asp Ile Gln Ser Ile Arg
30 35 40

```
gct gcc aaa tta gct cat ctg gag agc aca cat ggc gct ggt agg aaa      255
Ala Ala Lys Leu Ala His Leu Glu Ser Thr His Gly Ala Gly Arg Lys
    45                  50                  55

gag atg gac aac aat tta ggc agt tcc aat gag gac ata ttg cct tta      303
Glu Met Asp Asn Asn Leu Gly Ser Ser Asn Glu Asp Ile Leu Pro Leu
60                  65                  70                  75

aag aat tac ctg gat gcc cag tac tat gga gag att gga att ggt act      351
Lys Asn Tyr Leu Asp Ala Gln Tyr Tyr Gly Glu Ile Gly Ile Gly Thr
                80                  85                  90

cca cct cag aag ttc aca gtt ata ttt gat aca ggc agt tcc aac ctc      399
Pro Pro Gln Lys Phe Thr Val Ile Phe Asp Thr Gly Ser Ser Asn Leu
            95                  100                 105

tgg gtg cca tcg gca aaa tgt tac ttc tct att gcc tgc tgg ctc cac      447
Trp Val Pro Ser Ala Lys Cys Tyr Phe Ser Ile Ala Cys Trp Leu His
        110                 115                 120

tcc aag tac aag gca aag aag tca agt act tat aca gcc ata ggg aaa      495
Ser Lys Tyr Lys Ala Lys Lys Ser Ser Thr Tyr Thr Ala Ile Gly Lys
    125                 130                 135

tct tgt tca att cgt tat ggt tct gga tca att tct gga ttc tcc agt      543
Ser Cys Ser Ile Arg Tyr Gly Ser Gly Ser Ile Ser Gly Phe Ser Ser
140                 145                 150                 155

cag gat aac gtt gaa gtt ggt gat ctt gtt gtc aaa gat caa gtt ttt      591
Gln Asp Asn Val Glu Val Gly Asp Leu Val Val Lys Asp Gln Val Phe
                160                 165                 170

att gaa gct tca cga gaa gga agt ctt aca ttt gta att gcc aag ttt      639
Ile Glu Ala Ser Arg Glu Gly Ser Leu Thr Phe Val Ile Ala Lys Phe
            175                 180                 185

gac ggg ata ctt ggc ctt gga ttc cag gag atc gct gtt gat aac atg      687
Asp Gly Ile Leu Gly Leu Gly Phe Gln Glu Ile Ala Val Asp Asn Met
        190                 195                 200

gtg ccg gtc tgg tat aat atg gtg gac caa ggt ctc gtg gat gag caa      735
Val Pro Val Trp Tyr Asn Met Val Asp Gln Gly Leu Val Asp Glu Gln
    205                 210                 215

gta ttc tct ttc tgg ctt aac cgc gac cca aat gct gaa gac gga ggt      783
Val Phe Ser Phe Trp Leu Asn Arg Asp Pro Asn Ala Glu Asp Gly Gly
220                 225                 230                 235

gag ctg gtc ttt ggt ggt gta gat aca aat cac ttc aag gga aag cat      831
Glu Leu Val Phe Gly Gly Val Asp Thr Asn His Phe Lys Gly Lys His
                240                 245                 250

aca tat gtt cct gta act cag aag gga tac tgg caa ttt aaa atg gga      879
Thr Tyr Val Pro Val Thr Gln Lys Gly Tyr Trp Gln Phe Lys Met Gly
            255                 260                 265

gat ttt ctc att ggg aac gtc tca aca ggc ttt tgt gaa gga ggt tgt      927
Asp Phe Leu Ile Gly Asn Val Ser Thr Gly Phe Cys Glu Gly Gly Cys
        270                 275                 280

gct gct att gtg gac tct gga aca tcg ttg ctc gct ggt cca act act      975
Ala Ala Ile Val Asp Ser Gly Thr Ser Leu Leu Ala Gly Pro Thr Thr
    285                 290                 295

gtt gtg act caa att aat cat gcc att gga gct gaa gga gta gtt agc     1023
Val Val Thr Gln Ile Asn His Ala Ile Gly Ala Glu Gly Val Val Ser
300                 305                 310                 315

act gaa tgt aaa gaa att gtt tca cag tat ggt gaa ctg att tgg gat     1071
```

Thr Glu Cys Lys Glu Ile Val Ser Gln Tyr Gly Glu Leu Ile Trp Asp
320 325 330

ctc ctc gta tca ggg gta cta ccc gac aga gtt tgt aaa caa gct ggt 1119
Leu Leu Val Ser Gly Val Leu Pro Asp Arg Val Cys Lys Gln Ala Gly
335 340 345

tta tgt ccc ctt cgt ggt gct cag cat gag aat gct tat atc aag tca 1167
Leu Cys Pro Leu Arg Gly Ala Gln His Glu Asn Ala Tyr Ile Lys Ser
350 355 360

gtc gtc gac gag gag aac aag gag gaa gct tct gtt ggt gaa tcc ccg 1215
Val Val Asp Glu Glu Asn Lys Glu Glu Ala Ser Val Gly Glu Ser Pro
365 370 375

atg tgt act gct tgt gaa atg gct gtt gtt tgg atg caa aac cag ctg 1263
Met Cys Thr Ala Cys Glu Met Ala Val Val Trp Met Gln Asn Gln Leu
380 385 390 395

aaa cag cag gga act aag gag aaa gtg ctt gca tat gtg aat cag ctt 1311
Lys Gln Gln Gly Thr Lys Glu Lys Val Leu Ala Tyr Val Asn Gln Leu
400 405 410

tgt gaa agc ata cca agt ccc atg gga gaa tcc atc att gac tgc aac 1359
Cys Glu Ser Ile Pro Ser Pro Met Gly Glu Ser Ile Ile Asp Cys Asn
415 420 425

agt tta tcc acc ctg cca aat gtt tca ttc acc atc gga ggg aaa agt 1407
Ser Leu Ser Thr Leu Pro Asn Val Ser Phe Thr Ile Gly Gly Lys Ser
430 435 440

ttt gag ctg acc ctt aag gag tat gtt ctt cga act gga gaa ggc ttt 1455
Phe Glu Leu Thr Leu Lys Glu Tyr Val Leu Arg Thr Gly Glu Gly Phe
445 450 455

gct gaa gtc tgc atc agt gga ttc atg gct atg gat gtg ccg ccg cct 1503
Ala Glu Val Cys Ile Ser Gly Phe Met Ala Met Asp Val Pro Pro Pro
460 465 470 475

cgt ggt ccc atc tgg gtt ctg gga gat gtg ttc atg gga gtg tac cac 1551
Arg Gly Pro Ile Trp Val Leu Gly Asp Val Phe Met Gly Val Tyr His
480 485 490

acc gtg ttt gat tat ggt aat ctc cgg atg ggt ttc gca aga gct gct 1599
Thr Val Phe Asp Tyr Gly Asn Leu Arg Met Gly Phe Ala Arg Ala Ala
495 500 505

tag acaagactgt ttatttcgtc tactgtttga cggtcctaag agaagctatg 1652

aagacatgta gtagcttgta aattaggatt taattatgct tggctggttt atgggtggtg 1712

cttttaatat tatatgtaat gtaagcagat atgttacctt gttttagagt ttcaaggaaa 1772

ctgcaatatt tacttccggt aaaaaaaaaa aaaaaaaaaa aaaaaaa 1819

<210> 8

<211> 507

<212> PRT

<213> Coffea canephora

<400> 8

Met Glu Arg Arg Tyr Leu Trp Ala Ala Phe Val Leu Gly Ala Ile Val

1 5 10 15

Cys Ser Leu Phe Pro Leu Pro Ser Glu Gly Leu Lys Arg Ile Ser Leu
20 25 30

Lys Lys Lys Pro Leu Asp Ile Gln Ser Ile Arg Ala Ala Lys Leu Ala
35 40 45

His Leu Glu Ser Thr His Gly Ala Gly Arg Lys Glu Met Asp Asn Asn
50 55 60

Leu Gly Ser Ser Asn Glu Asp Ile Leu Pro Leu Lys Asn Tyr Leu Asp
65 70 75 80

Ala Gln Tyr Tyr Gly Glu Ile Gly Ile Gly Thr Pro Pro Gln Lys Phe
85 90 95

Thr Val Ile Phe Asp Thr Gly Ser Ser Asn Leu Trp Val Pro Ser Ala
100 105 110

Lys Cys Tyr Phe Ser Ile Ala Cys Trp Leu His Ser Lys Tyr Lys Ala
115 120 125

Lys Lys Ser Ser Thr Tyr Thr Ala Ile Gly Lys Ser Cys Ser Ile Arg
130 135 140

Tyr Gly Ser Gly Ser Ile Ser Gly Phe Ser Ser Gln Asp Asn Val Glu
145 150 155 160

Val Gly Asp Leu Val Val Lys Asp Gln Val Phe Ile Glu Ala Ser Arg
165 170 175

Glu Gly Ser Leu Thr Phe Val Ile Ala Lys Phe Asp Gly Ile Leu Gly
180 185 190

Leu Gly Phe Gln Glu Ile Ala Val Asp Asn Met Val Pro Val Trp Tyr
195 200 205

Asn Met Val Asp Gln Gly Leu Val Asp Glu Gln Val Phe Ser Phe Trp
210 215 220

Leu Asn Arg Asp Pro Asn Ala Glu Asp Gly Gly Glu Leu Val Phe Gly
225 230 235 240

Gly Val Asp Thr Asn His Phe Lys Gly Lys His Thr Tyr Val Pro Val
245 250 255

Thr Gln Lys Gly Tyr Trp Gln Phe Lys Met Gly Asp Phe Leu Ile Gly
260 265 270

Asn Val Ser Thr Gly Phe Cys Glu Gly Gly Cys Ala Ala Ile Val Asp
275 280 285

Ser Gly Thr Ser Leu Leu Ala Gly Pro Thr Thr Val Val Thr Gln Ile
290 295 300

Asn His Ala Ile Gly Ala Glu Gly Val Val Ser Thr Glu Cys Lys Glu
305 310 315 320

Ile Val Ser Gln Tyr Gly Glu Leu Ile Trp Asp Leu Leu Val Ser Gly
325 330 335

Val Leu Pro Asp Arg Val Cys Lys Gln Ala Gly Leu Cys Pro Leu Arg
340 345 350

Gly Ala Gln His Glu Asn Ala Tyr Ile Lys Ser Val Val Asp Glu Glu
355 360 365

Asn Lys Glu Glu Ala Ser Val Gly Glu Ser Pro Met Cys Thr Ala Cys
370 375 380

Glu Met Ala Val Val Trp Met Gln Asn Gln Leu Lys Gln Gln Gly Thr
385 390 395 400

Lys Glu Lys Val Leu Ala Tyr Val Asn Gln Leu Cys Glu Ser Ile Pro
405 410 415

Ser Pro Met Gly Glu Ser Ile Ile Asp Cys Asn Ser Leu Ser Thr Leu
420 425 430

Pro Asn Val Ser Phe Thr Ile Gly Gly Lys Ser Phe Glu Leu Thr Leu
435 440 445

Lys Glu Tyr Val Leu Arg Thr Gly Glu Gly Phe Ala Glu Val Cys Ile
450 455 460

Ser Gly Phe Met Ala Met Asp Val Pro Pro Pro Arg Gly Pro Ile Trp
465 470 475 480

Val Leu Gly Asp Val Phe Met Gly Val Tyr His Thr Val Phe Asp Tyr
485 490 495

Gly Asn Leu Arg Met Gly Phe Ala Arg Ala Ala
500 505